Improved Strength Four Covering Arrays with Three Symbols

Soumen Maity^{*a*} Yasmeen Akhtar^{*a*}

Reshma C. Chandrasekharan^{*a*} Charles J. Colbourn^{*b*}

^aIndian Institute of Science Education and Research Pune, India ^bSchool of Computing, Informatics and Decision Systems Engineering,

Arizona State University, U.S.A

Abstract

A covering array t-CA(n,k,g), of size n, strength t, degree k, and order g, is a $k \times n$ array on g symbols such that every $t \times n$ sub-array contains every $t \times 1$ column on g symbols at least once. Covering arrays have been studied for their applications to software testing, hardware testing, drug screening, and in areas where interactions of multiple parameters are to be tested. In this paper, we present an algebraic construction that improves many of the best known upper bounds on n for covering arrays 4-CA(n,k,g) with g = 3. The coverage measure $\mu_t(A)$ of a testing array A is defined by the ratio between the number of distinct t-tuples contained in the column vectors of A and the total number of t-tuples. A covering array is a testing array with full coverage. The covering arrays with budget constraints problem is the problem of constructing a testing array of size at most *n* having largest possible coverage measure, given values of k, g and n. This paper presents several strength four testing arrays with high coverage. The construction here is a generalisation of the construction methods used by Chateauneuf, Colbourn and Kreher, and Meagher and Stevens.

1 Introduction

This article focuses on constructing new strength-four covering arrays with g = 3and establishing improved bounds on the covering array numbers 4-CAN(k,3). This article also presents solution to the covering arrays with budget constraints problem by constructing many strength four testing arrays with high coverage. A covering array t-CA(n,k,g), of size n, strength t, degree k, and order g, is a $k \times n$ array on g symbols such that every $t \times n$ sub-array contains every $t \times 1$ column on g symbols at least once. It is desirable in most applications to minimise the size n of covering arrays. The covering array number t-CAN(k,g) is the smallest n for which a t-CA(n,k,g) exists. An obvious lower bound is

$$g^t \leq t$$
-CAN (k,g)

In this paper, we describe a construction method which is an extension of the methods developed by Chateauneuf, Colbourn and Kreher [1] and Meagher and Stevens [13]. This method improves some of the best known upper bounds for strength four covering arrays with g = 3. In the range of degrees considered in this paper, the best known results previously come from [5]; in that paper, covering arrays are also found by using a group action on the symbols (the affine or Frobenius group), but no group action on the rows is employed. While for g = 3 the group that we employ on the symbols coincides with the affine group, we accelerate and improve the search by also exploiting a group action on the rows as in [1, 13], and develop a search method than can be applied effectively whenever $g \ge 3$ and g - 1 is a prime power.

There is a large literature [1, 7] on covering arrays, and the problem of determining small covering arrays has been studied under many guises over the past thirty years. In [7], Hartman and Raskin discussed several generalizations motivated by their applications in the realm of software testing. When testing a software system with k parameters, each of which must be tested with g values, the total number of possible test cases is g^k . For instance, if there are 20 parameters and three values for each parameter then the number of input combinations or test cases of this system is $3^{20} = 3486784401$. A fundamental problem with software testing is that testing under all combinations of inputs is not feasible, even with a simple product [9, 10]. Software developers cannot test everything, but they can use combinatorial test design to identify the minimum number of tests needed to get the coverage they want. The goal of most combinatorial testing research is to create test suites that find a large percentage of errors of a system while having a small number of tests required. Covering arrays prove useful in locating a large percentage of errors in software systems [3, 16]. The test cases are the columns of a covering array t-CA(n,k,g). This is one of the five natural generalizations in [7]. Covering arrays with budget constraints: A practical limitation in the realm of testing is budget. In most software development environments, time, computing, and human resources needed to perform the testing of a component is strictly limited. To model this situation, we consider the problem of creating best possible test suite (covering the maximum number of *t*-tuples) within a fixed number of test cases. The coverage measure $\mu_t(A)$ of a testing array A is defined by the ratio between the number of distinct t-tuples contained in the column vectors of A and the total number of t-tuples given by $\binom{k}{t}g^t$. Our objective is to construct a testing array A of size at most *n* having largest possible coverage measure, given fixed values of t, k, g and n. This problem is called *covering arrays with budget constraints*.

We summarize the results from group theory that we use. Let \mathbb{F}_q be a Galois field GF(q) where $q = p^m$ and p is prime. We adjoin to \mathbb{F}_q the symbol ∞ : it may be helpful to think of the resulting set

$$X = \mathbb{F}_q \cup \{\infty\}$$

as the projective line consisting of q + 1 points. Recall that the projective general

linear group of dimension 2 may be seen as the "fractional linear group":

$$PGL(2,q) = \{ \alpha : X \mapsto X \mid x\alpha = \frac{ax+b}{cx+d}, \text{ where } a, b, c, d \in \mathbb{F}_q \text{ and } ad-bc \neq 0 \}$$

in which we define $\frac{1}{0} = \infty$, $\frac{1}{\infty} = 0$, $1 - \infty = \infty - 1 = \infty$, and $\frac{\infty}{\infty} = 1$. It is known that $|PGL(2,q)| = \frac{(q^2-1)(q^2-q)}{(q-1)} = (q+1)q(q-1)$ and its action on $\mathbb{F}_q \cup \{\infty\}$ is sharply 3-transitive. For the undefined terms and more details see [15, Chapter 7].

Pair-wise or 2-way interaction testing and 3-way interaction testing are known to be effective for different types of software testing [3, 11, 12]. However, software failures may be caused by interactions of more than three parameters. A recent NIST study indicates that failures can be triggered by interactions up to 6 parameters [10]. Here we consider the problem of 4-way interaction testing of the parameters. The construction given in this paper improves many of the current best known upper bounds on 4-*CAN*(k,g) with g = 3 and $21 \le k \le 74$. This paper also presents several strength four testing arrays with high coverage measures.

2 PGL Construction

Let $X = GF(g-1) \cup \{\infty\}$ be the set of *g* symbols on which we are to construct a 4-*CA*(*n*,*k*,*g*). We choose *g* so that g-1 is a prime or prime power.

2.1 Case 1: Two starter vectors

Our construction involves selecting a group *G* and finding vectors $u, v \in X^k$, called starter vectors. We use the vectors to form a $k \times 2k$ matrix *M*.

$$M = \begin{pmatrix} u_1 & u_k & \dots & u_2 & v_1 & v_k & \dots & v_2 \\ u_2 & u_1 & \dots & u_3 & v_2 & v_1 & \dots & v_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{k-1} & u_{k-2} & \dots & u_k & v_{k-1} & v_{k-2} & \dots & v_k \\ u_k & u_{k-1} & \dots & u_1 & v_k & v_{k-1} & \dots & v_1 \end{pmatrix}.$$

Let G = PGL(2, g - 1). For each $a \in PGL(2, g - 1)$, let M^a be the matrix formed by the action of a on the elements of M. The matrix obtained by developing M by G is the $k \times 2k|G|$ matrix $M^G = [M^a : a \in G]$. Let C be the $k \times g$ matrix that has a constant column with each entry equal to x, for each $x \in X$. Vectors $u, v \in X^k$ are said to be *starter vectors* for a 4-CA(n,k,g) if any $4 \times 2k$ subarray of the matrix M has at least one representative from each non-constant orbit of PGL(2,g-1)acting on 4-tuples from X. Under this group action, there are precisely g + 11orbits of 4-tuples. These g + 11 orbits are determined by the pattern of entries in their 4-tuples:

1.
$$\{[a, a, a, a]^T : a \in X\}$$

- 2. $\{[a,a,a,b]^T : a,b \in X, a \neq b\}$ 3. $\{[a,a,b,a]^T : a,b \in X, a \neq b\}$ 4. $\{[a,b,a,a]^T : a,b \in X, a \neq b\}$ 5. $\{[b,a,a,a]^T : a,b \in X, a \neq b\}$ 6. $\{[a,a,b,b]^T : a,b \in X, a \neq b\}$ 7. $\{[a,b,a,b]^T : a,b \in X, a \neq b\}$ 8. $\{[a,b,b,a]^T : a,b \in X, a \neq b\}$ 9. $\{[a,a,b,c]^T : a,b,c \in X, a \neq b \neq c\}$ 10. $\{[b,a,a,c]^T : a,b,c \in X, a \neq b \neq c\}$ 11. $\{[a,b,a,c]^T : a,b,c \in X, a \neq b \neq c\}$ 12. $\{[b,a,c,a]^T : a,b,c \in X, a \neq b \neq c\}$ 13. $\{[a,b,c,a]^T : a,b,c \in X, a \neq b \neq c\}$ 14. $\{[b,c,a,a]^T : a,b,c \in X, a \neq b \neq c\}$
- 15. g-3 orbits of patterns with four distinct entries. The reason is this. There are g(g-1)(g-2)(g-3) 4-tuples with four distinct entries and each orbit contains g(g-1)(g-2) 4-tuples as |PGL(2,g-1)| = g(g-1)(g-2).

If starter vectors u, v exist in X^k (with respect to the group G) then there exists a 4-CA(2kg(g-1)(g-2)+g,k,g). We give an example to explain the method.

Example 1. Let g = 3, k = 30, $X = GF(2) \cup \{\infty\}$ and G = PGL(2,2). The action of *G* on 4-tuples from *X* has 14 orbits:

- Orb 1: [0000,∞∞∞∞,1111]
- Orb 2: [0001,000∞,∞∞∞0,∞∞∞1,1110,111∞]
- Orb 3: [1∞∞∞, 1000, 0111, ∞000, 0∞∞∞, ∞111]
- Orb 4: $[0100, \infty0\infty\infty, 0\infty00, \infty1\infty\infty, 1011, 1\infty11]$
- Orb 5: $[11\infty1, \infty\infty1\infty, 0010, 1101, 00\infty0, \infty\infty0\infty]$
- Orb 6: [11∞∞,∞∞11,0011,1100,00∞∞,∞∞00]
- Orb 7: $[\infty 0 \infty 0, 0101, \infty 1 \infty 1, 0 \infty 0 \infty, 1010, 1 \infty 1 \infty]$
- Orb 8: [\infty11\infty, 1\infty\infty1, 1001, 0110, \infty00\infty, 0\infty\infty]

Orb 9: $[11\infty0, \infty\infty10, 001\infty, 110\infty, 00\infty1, \infty\infty01]$

Orb 10: $[\infty 0 \infty 1, 010 \infty, \infty 1 \infty 0, 0 \infty 01, 101 \infty, 1 \infty 10]$

Orb 11: $[1 \approx 01, 0 \approx 10, \infty 10 \approx, 01 \approx 0, \infty 01 \approx, 10 \approx 1]$

Orb 12: $[1 \approx 0 \approx, 0 \approx 1 \approx, \infty 101, 01 \approx 1, \infty 010, 10 \approx 0]$

Orb 13: $[1 \approx 00, 0 \approx 11, \infty 100, 01 \approx \infty, \infty 011, 10 \approx \infty]$

Orb 14:
$$[1\infty\infty0, 100\infty, 011\infty, \infty001, 0\infty\infty1, \infty110]$$

The following are starter vectors to construct $[M^G, C]$, a 4-CA(363, 30, 3):

We used computer search to find *u* and *v*. One can check that on each set of 4 rows of *M* there is a representative from each orbit 2 - 14. Thus, $4-CAN(30,3) \le 363$.

2.2 Choice of starter vectors *u* and *v*

The problem is to find two vectors $u, v \in X^k$ such that on each set of 4 rows of M there is a representative from each orbit 2-15. To determine which vectors work as starters, we define the sets d[x, y, z] for positive integers x, y and z as follows:

$$d[x,y,z] = \{(u_i, u_{i+x}, u_{i+x+y}, u_{i+x+y+z}) : 0 \le i \le k-1\} \bigcup_{\{(v_i, v_{i+x}, v_{i+x+y}, v_{i+x+y+z}) : 0 \le i \le k-1\}}$$

where the subscripts are taken modulo k. For computational convenience, we partition the collection of $\binom{k}{4}$ choices of four distinct rows from k rows into disjoint equivalence classes.

Formally, let *S* be the set of all $\binom{k}{4}$ 4-combinations of the set $\{1, 2, ..., k\}$. Define a binary relation *R* on *S* by putting

$$\{s_1, s_2, s_3, s_4\} R \{s'_1, s'_2, s'_3, s'_4\}$$
iff
$$\{s_1 + d, s_2 + d, s_3 + d, s_4 + d\} = \{s'_1, s'_2, s'_3, s'_4\}$$
for some $d \in \mathbb{N}$

where all of the addition is modulo k. Because R is an equivalence relation on S, S can be partitioned into disjoint equivalence classes. The equivalence class determined by $\{s_1, s_2, s_3, s_4\} \in S$ is given by

$$[\{s_1, s_2, s_3, s_4\}] = \{\{s_1 + d, s_2 + d, s_3 + d, s_4 + d\} | 0 \le d \le k - 1\}.$$

Without loss of generality, we may assume that $0 = s_1 < s_2 < s_3 < s_4$ for each equivalence class representative $[\{s_1, s_2, s_3, s_4\}]$. As an illustration, when $X = \{0, 1, 2, ..., 7\}$. *S* is partitioned into 10 disjoint equivalence classes:

$$[\{0,1,2,3\}] \quad [\{0,1,2,4\}] \quad [\{0,1,2,5\}] \quad [\{0,1,2,6\}] \quad [\{0,1,3,4\}]$$

 $[\{0,1,3,5\}] \quad [\{0,1,3,6\}] \quad [\{0,1,4,5\}] \quad [\{0,1,4,6\}] \quad [\{0,2,4,6\}]$

A distance vector (x, y, z, w) is associated with every equivalence class $[\{s_1, s_2, s_3, s_4\}]$ where $x = s_2 - s_1$, $y = s_3 - s_2$, $z = s_4 - s_3$, $w = s_1 - s_4 \mod k$. The fourth distance is redundant because x + y + z + w = k. We rewrite the equivalence class of 4combinations $[\{s_1, s_2, s_3, s_4\}]$ as

$$[x, y, z] = \{i, i+x, i+x+y, i+x+y+z\} | i = 0, 1, 2, \dots, k-1\}$$

For k = 8, $[1,1,1] = [\{0,1,2,3\}]$, $[1,1,2] = [\{0,1,2,4\}]$, $[1,1,3] = [\{0,1,2,5\}]$, $[1,1,4] = [\{0,1,2,6\}]$, $[1,2,1] = [\{0,1,3,4\}]$, $[1,2,2] = [\{0,1,3,5\}]$, $[1,2,3] = [\{0,1,3,6\}]$, $[1,3,1] = [\{0,1,4,5\}]$, $[1,3,2] = [\{0,1,4,6\}]$, $[2,2,2] = [\{0,2,4,6\}]$.

Lemma 1. Let *S* be the set of all 4-combinations of $\{1, 2, 3, ..., k\}$. Then *S* can be partitioned into disjoint equivalence classes

$$[x, y, z] = \{i, i+x, i+x+y, i+x+y+z\} | i = 0, 1, 2, \dots, k-1\}$$

where $x = 1, 2, ..., \lfloor \frac{k}{4} \rfloor$, y = x, x + 1, ..., k - 1 and z = x, x + 1, ..., k - 1 such that

- (*i*) 2x + y + z < k
- (*ii*) when x = z, $x \le y \le \lfloor \frac{k-2x}{2} \rfloor$

There are no further classes distinct from these.

Before proving the result, we give an example. When *S* is the set of all 4-combinations of $\{0, 1, 2, 3, 4, 5, 6, 7\}$, *S* can be partitioned into 10 disjoint classes: [1, 1, 1], [1, 1, 2], [1, 1, 3], [1, 1, 4], [1, 2, 1], [1, 2, 2], [1, 3, 1], [1, 3, 2] and [2, 2, 2].

Proof. Let (x, y, z, w) be the distance vector corresponding to equivalence class $[\{s_1, s_2, s_3, s_4\}]$. Classes $[\{s_1, s_2, s_3, s_4\}]$, [x, y, z], [y, z, w], [z, w, x] and [w, x, y] are the same. Without loss of generality, we choose [x, y, z] as class representative if $x \le y, x \le z$. Thus $1 \le x \le \frac{k}{4}, y = x, x + 1, ..., k - 1$ and z = x, x + 1, ..., k - 1. We consider three cases: (i) x = w, (ii) x = z, (iii) x = y. If w = x, then the classes [x, y, z] and [x, x, y] obtained from distance vector (x, y, z, x) are the same equivalence class. The classes of the form [x, x, y] are generated under case (iii) as well. In order to avoid repetition, w has to be strictly greater than x. That is, w = k - x - y - z > x which implies 2x + y + z < k. If z = x, then the classes [x, y, z] and [x, w, x] are the same where y + w = k - 2x. Thus it is sufficient to consider the classes of the form [x, y, x] for $y \le \lfloor \frac{k-2x}{2} \rfloor$ only. Hence the lemma follows.

All the equivalence classes are enumerated by the following algorithm.

EQUIVALENCE-CLASSES(k) Input: k Output: All [x, y, z] classes. for $\mathbf{x} \leftarrow 1$ to $\frac{k}{4}$ do for $\mathbf{y} \leftarrow x$ to k-1 do

```
if y > \frac{k-2x}{2} then

for z \leftarrow x + 1 to k - 2x - y - 1 do

add [x, y, z]

end for

else

if y == \frac{k-2x}{2} and x == \frac{k-2x}{2} then

add [\frac{k}{4}, \frac{k}{4}, \frac{k}{4}]

else

for z \leftarrow x to k - 2x - y - 1 do

add [x, y, z]

end for

end if

end if

end for

end for
```

Theorem 1. Let $X = GF(g-1) \cup \{\infty\}$ and G = PGL(2, g-1). If there exists a pair of vectors $u, v \in X^k$ such that each d[x, y, z] has a representative from each of the orbits 2-15, then there exists a 4-CA(2kg(g-1)(g-2)+g, k, g) covering array.

Proof. Let $u, v \in X^k$ be vectors such that each d[x, y, z] has a representation from each of the orbits 2-15. Using u, v, we create the matrix $[M^G, C]$. Let $\{s_1, s_2, s_3, s_4\}$ be a member in *S*. By Lemma 1, there exists three positive integers x_0 , y_0 and z_0 such that $\{s_1, s_2, s_3, s_4\} \in [x_0, y_0, z_0]$. It is given that $d[x_0, y_0, z_0]$ has a representative from each of the orbits 2-15. In other words, if we look at the rows s_1, s_2, s_3, s_4 of *M*, we see representative from each of the g + 11 orbits. Consequently, because PGL(2, g - 1) is 3-transitive on *X*, $[M^G, C]$ is a 4-CA(2kg(g - 1)(g - 2) + g, k, g).

At this stage, we make a few remarks about the size of equivalence classes defined by above choices of x, y and z.

1. $k \not\equiv 0 \mod 2$:

If *k* is an odd integer, each class contains exactly *k* distinct choices from the collection of $\binom{k}{4}$ choices and hence there are $l = \frac{(k-1)(k-2)(k-3)}{24}$ distinct classes of size *k*.

2. $k \equiv 0 \mod 2$:

If k is an even integer, $\frac{k}{2}$ can be written as sum of two positive integers a and

b where $a \le b$ in $\lfloor \frac{k}{4} \rfloor$ different ways.

Case 1 : If $k \neq 0 \mod 4$, a class of the form [a, b, a] contains only $\frac{k}{2}$ distinct choices. There are total $\lfloor \frac{k}{4} \rfloor$ equivalence classes of the form [a, b, a] with size $\frac{k}{2}$ and the remaining classes are of size *k*.

Case 2 : If $k \equiv 0 \mod 4$, a class of the form [a, b, a] contains only $\frac{k}{2}$ distinct choices and a class of the form [a, a, a] where $a = \frac{k}{4}$ contains only $\frac{k}{4}$ distinct choices. Here we get total $\frac{k}{4} - 1$ equivalence classes of size $\frac{k}{2}$, exactly one class of size $\frac{k}{4}$ and the remaining classes are of size k.

For k = 8, there are 10 equivalence classes. The classes [1,3,1] and [2,2,2] are of size 4 and 4 respectively and the remaining 8 classes are of size 8 each. Thus $8 \times 8 + 4 + 2 = \binom{8}{4}$.

2.3 Case 2: Two vectors u, v and a matrix C_1

If we do not find vectors u and v such that each d[x, y, z] contains a representative from each of the orbits 2 - 15, we look for vectors that produce an array with maximum possible coverage. In order to complete the covering conditions, we add a small matrix C_1 . We give an example below to illustrate the technique.

Example 2. Let k = 21 and g = 3. Here we do not find vectors u and v such that each d[x, y, z] contains a representative from each of the orbits 2 - 15. For k = 21, there are 285 [x, y, z] classes. All classes [x, y, z] are obtained by the algorithm EQUIVALENCE-CLASSES. One can check that for the vectors

 $u = 00001010 \times 10000001 \times 1$ $v = 0000100 \times 00001001 \times 11100$

there is a representative from each orbit 2-15 on 276 of the d[x, y, z] classes. Table 1 shows nine classes which do not have representative from all the orbits:

Class	Missing orbits
d[1,2,2]	10
d[1,5,6]	2
d[1, 6, 12]	5
d[1, 13, 5]	9
d[2,3,8]	6
d[2,7,3]	10
d[2, 12, 3]	13
d[3, 6, 8]	6
d[3,7,7]	10

Table 1: List of classes not having representative from all the orbits

In order to complete the covering conditions, we add a small matrix C_1 .

$$C_{1} = \begin{pmatrix} \infty & 0 & 1 & 1 & 0 & \infty & \infty & \infty & 1 \\ \infty & 1 & 1 & \infty & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & \infty & 1 & 1 & 0 & 1 & \infty & 0 \\ 0 & \infty & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & \infty & 0 & \infty & \infty & \infty & 1 \\ \infty & \infty & \infty & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & \infty & 1 & \infty & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & \infty & 1 & \infty & 0 \\ \infty & \infty & \infty & \infty & 0 & 0 & 1 & \infty & 0 \\ \infty & 1 & \infty & 1 & 0 & 1 & \infty & \infty & \infty \\ 0 & 1 & 1 & 0 & \infty & 1 & \infty & 1 & 0 \\ 0 & \infty & 1 & 0 & \infty & \infty & 0 & 0 & 0 \\ 0 & 0 & 0 & \infty & 1 & \infty & 1 & 0 & 0 \\ \infty & 0 & 0 & 1 & \infty & 0 & 0 & 0 \\ \infty & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & \infty & \infty & \infty & \infty & \infty & 1 & 0 & 0 \\ \infty & 0 & 0 & 1 & \infty & 0 & \infty & \infty & 1 \end{pmatrix}$$

We use computer search to find matrix C_1 . This matrix has the property that every choice of four rows in [1,2,2], [2,7,3] and [3,7,7] contains at least one representative from orbit 10; every choice of four rows in [2,3,8] and [3,6,8] contains at least one representative from orbit 6; each choice of four rows in [1,5,6], [1,6,12], [1,13,5] and [2,12,3] contains at least one representative from orbit 2, 5, 9 and 13 respectively. We also need to use the following matrix

$$C = \begin{pmatrix} 0 & 1 & \infty \\ 0 & 1 & \infty \\ \vdots & \vdots & \vdots \\ 0 & 1 & \infty \end{pmatrix}$$

to ensure the coverage of all identical 4-tuples. Therefore, $[M^G, C_1^G, C]$ is a 4-CA(315, 21, 3).

2.4 Case 3: One vector u and a matrix C_1

For k = 37 to 58, we use one starter vector and a C_1 matrix of order $k \times \ell$ with $\ell < k$. Tables 2, 3, 4 and 5 give a list of starter vectors and matrix C_1 that improves the best known bounds. When the new bound is marked with an asterisk, post-optimization has been applied (see Section 3.2).

3 Improving the solutions

We examine two methods to obtain small improvements on the computational results obtained.

3.1 Extending a solution

Until this point, starter vectors have been developed by applying a cyclic rotation of the starter vectors in addition to the action of PGL on the symbols. As in [13], one can also consider fixing one row, and developing the remaining k-1 cyclically. This can be viewed as first finding a solution of the type already described on k-1 rows, but requiring an additional property. For the 4-subsets of $\{0, \ldots, k-2\}$, equivalence classes are defined as before, with arithmetic modulo k-1:

$$[\{s_1, s_2, s_3, s_4\}] = \{\{s_1 + d, s_2 + d, s_3 + d, s_4 + d\} | 0 \le d \le k - 2\}$$

For 3-subsets $\{t_1, t_2, t_3\}$ of $\{0, \dots, k-2\}$ we define further equivalence classes as

$$[\{t_1, t_2, t_3, k-1\}] = \{\{t_1 + d, t_2 + d, t_3 + d, k-1\} | 0 \le d \le k-2\}.$$

If we can place an entry in position k - 1 to extend the length of each starter vector so that every one of the (old and new) equivalence classes represents each of the orbits 2 - 15, we obtain a 4-CA of degree k.

The potential advantage of this approach is that a solution for degree k - 1 can sometimes be extended to one of degree k without increasing the size of the covering array produced. Indeed we found that the solutions for $k - 1 \in \{32, 34, 35\}$ do ensure that the new equivalence classes also represent each of the orbits 2 - 15. Hence we obtain the following improvements. Old indicates the bound obtained by applying our methods to k; Improved gives the bound by applying the method to k - 1 and ensuring that the new equivalence classes represent all orbits:

k	Old	Improved	k	Old	Improved	k	Old	Improved
33	399	387	35	423	411	36	435	423

3.2 Randomized Post-optimization

Nayeri, Colbourn, and Konjevod [14] describe a post-optimization strategy which, when applied to a covering array, exploits flexibility of symbols in an attempt to reduce its size. We applied their method to the arrays provided here, and to arrays obtained by removing one or more rows. Because the method is described in detail elsewhere, we simply report improvements for eight values of k. Basic gives the bound from starter vectors, Improved gives the bound on 4-CAN(k,3) after post-optimization:

k	Basic	Improved	k	Basic	Improved	k	Basic	Improved
19	309	300	20	309	303	21	309	305
22	309	307	27	351	345	28	363	360
34	411	410	37	435	433			

4 Covering arrays with budget constraints problem

In this section we present several strength four testing arrays with high coverage measure for $g \ge 3$. The coverage measure $\mu_4(A)$ of a strength four testing array A is defined by the ratio between the number of distinct 4-tuples contained in the column vectors of A and the total number of 4-tuples given by $\binom{k}{4}g^4$. Note that the coverage measure of a covering array is always one. For computational convenience, we rewrite the coverage measure in terms of equivalence classes [x, y, z] and d[x, y, z] as follows:

$$\mu_4(A) = \frac{\sum_{x,y,z} |[x,y,z]| \times \text{number of distinct 4-tuples covered by } d[x,y,z]}{\binom{k}{4}g^4}.$$

We search by computer to find vectors v with very high coverage measures. Tables 6 and 7 show vectors with high coverage, the number of test cases (n) generated by our technique, and the best known size with full coverage. Comparison of our construction with best known covering array sizes shows that our construction produces significantly smaller testing arrays with very high coverage measures.

5 Conclusions

In this paper, we present a construction method of strength four covering arrays with three symbols that combines an algebraic technique with computer search. This method improves the current best known upper bounds on 4-*CAN*(k,g) for $21 \le k \le 74$ and g = 3. We have also proposed a construction of strength four covering arrays with budget constraints. In order to test software with 25 parameters each having three values, our construction can generate a test suite with 153 test cases that ensure with probability 0.93 that software failure cannot be caused due to interactions of two, three or four parameters whereas the best known covering array in [4] requires 363 test cases for full coverage. The results show that the proposed method could reduce the number of test cases significantly while compromising only slightly on the coverage.

Acknowledgements

The second author gratefully acknowledges support from the Council of Scientific and Industrial Research (CSIR), India, during the work under CSIR senior research fellow scheme. The fourth author's research was supported in part by the National Science Foundation under Grant No. 1421058.

References

References

- [1] M.A. Chateauneuf, C.J. Colbourn, D.L. Kreher, Covering Arrays of Strength Three. Designs, Codes and Cryptography, 16 (1999) 235-242.
- [2] M.A. Chateauneuf, D.L. Kreher, On the state of strength-three covering arrays, J. Combin. Design, 10(4) (2002) 217-238.
- [3] D.M. Cohen, S.R. Dalal, M.L. Fredman and G.C. Patton, The AETG system: An Approach to Testing Based on Combinatorial Design. IEEE Transaction on Software Engineering, 23 (7) (1997) 437-443.
- [4] C.J. Colbourn, Covering Array Tables for t=2,3,4,5,6 available at http://www.public.asu.edu/~ccolbou/src/tabby/catable.html
- [5] C.J. Colbourn, Conditional expectation algorithms for covering arrays, Journal of Combinatorial Mathematics and Combinatorial Computing, 90 (2014) 97-115.
- [6] Covering Arrays generated by IPOG-F, available at http://math.nist.gov/coveringarrays/ipof/ipof-results.html
- [7] A. Hartman, L. Raskin, Problems and algorithms for covering arrays. Discrete Mathematics, 284 (2004) 149-156.
- [8] A. Hartman, Software and hardware testing using combinatorial covering suites. Graph Theory, Combinatorics and Algorithms: Interdisciplinary Applications, Kluwer Academic Publishers, 34 (2006) 237-266.
- [9] C. Kaner, J. Falk, H.Q. Nguyen, Testing Computer Software. 2nd Ed., John Wiley and Sons, Inc., New York (1999).
- [10] D.R. Kuhn, D.R. Wallace, A.M. Gallo, Software Fault Interactions and Implications for Software Testing. IEEE Transactions on Software Engineering. 30 (6) (2004) 418-421.
- [11] S. Maity, 3-Way software testing with budget constraints. IEICE Transactions on Information and Systems, E-95-D, No.9 (2012) 2227-2231.
- [12] S. Maity, A. Nayak, Improved Test Generation Algorithms for Pair-Wise Testing. Proc. 16th IEEE International Symposium on Software Reliability Engineering, Chicago, pp. 235–244 (2005).
- [13] K. Meagher and B. Stevens, Group Construction of Covering Arrays. Journal of Combinatorial Designs. 13(1) (2005) 70-77.

- [14] P. Nayeri, C.J. Colbourn and G. Konjevod, Randomized post-optimization of covering arrays, Eur. J. Combin., 34(1) (2013) 91-103.
- [15] D.J.S. Robinson, A course in the theory of groups, Second Edition, Springer (1995).
- [16] C. Yilmaz, M. Cohen, A. Porter, Covering arrays for efficient fault characterisation in complex configuration spaces. IEEE Trans. on Software Engineering. 32(1) (2006) 20-34.

k	Starter vectors and matrix C_1	New	Old
		bound	bound
21	$u = (00001010 \times 10 \times 010 \times 001 \times 1)$ $v = (0000100 \times 000 \times 10001 \times 111 \times)$ $C_{1} = \begin{pmatrix} \infty & \infty & 0 & 0 & 1 & \infty & 0 & 0 & 1 & \infty & 0 & 0 & 0 & \infty & 1 & 0 & 0 & \infty \\ 0 & 1 & 1 & \infty & 0 & 0 & \infty & 1 & 1 & 0 & 0 & 0 & 1 & 1 & \infty & 0 \\ 1 & 1 & 0 & 0 & 0 & \infty & 1 & 0 & 0 & \infty & 1 & 1 & 0 & 0 & 0 & 1 & \infty & 0 \\ 1 & \infty & 1 & 0 & 0 & \infty & 1 & 1 & 0 & 0 & \infty & 1 & 0 & 1 & 0 & \infty & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & \infty & 1 & 1 & 0 & 0 & \infty & 1 & 0 & 0 & \infty & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0$	305*	315
22	$u = (0000011 \times 0 \times 0110 \times 1 \times 0 \times 01 \times 0)$ $v = (00010010 \times 1 \times 0 \times 0 \times 0 \times 0 \times 0 \times 0 \times 0$	307*	315
27	$u = (1101011 \times 00000 \times 0000001 \times 0011 \times 001000)$ $v = (11 \times 0001011 \times 00000000000000000000000$	345*	378

Table 2: Improved strength four covering arrays for g = 3.

ĸ	Starter vectors and matrix C_1	New bound	Old bound
28	$u = (1 \infty 00 0 \infty 1 01101111 00 00101 0 0001)$ $v = (\infty 1011 010 000 0 0 0 0 0 0 00000000000$	360*	383
30	$u = (011 \times 11 \times 0000 \times 001 \times 0000 \times 10000 \times 00001)$ $v = (11 \times 00000 \times 101000 \times 1000000000000000$	363	393
32	$u = (\infty 1100010 \infty 111 \infty 1 \infty 010 \infty 0100 \infty 00000000$	387	409
33	Obtained from <i>CA</i> (387, 32, 3)	387	417
34	$u = (00 \times 101 \times 001 \times 010 \times 000 \times 00 \times 0000 \times 0000 \times 0000)$ $v = (1100 \times 1001 \times 10110 \times 0000)$	410*	423
35	Obtained from $CA(411, 34, 3)$	411	429
35	$u = 01 \infty 0 \infty \infty 1000 \infty 01 \infty \infty 0 \infty 1 \infty 111 \infty \infty \infty 01 \infty 01000 \infty 1$ $v = 0 \infty 00111 \infty 0 \infty 110 \infty 11 \infty 110 \infty 01001000 0 \infty 1 \infty \infty 0$	423	429

Table 3: **Improved strength four covering arrays for** g = 3 (**continued**).

Table 4: Improved strength four covering arrays for $g = 3$ (continued).	
--	--

k	Starter vectors and matrix C_1	New bound	Old bound
36	Obtained from $CA(423, 35, 3)$	423	441
37	$u = (001 \approx 10 \approx 1 \approx 01000 \approx 1100 \approx 101111 \approx 001 \approx \infty \approx 00\infty)$ C ₁ : 37 × 35 matrix	433*	441
39	$u = (001\infty\infty11\infty11\infty0001\infty11\infty101\infty\infty\infty1\infty0\infty0010\infty000\infty00)$ C ₁ : 39 × 34 matrix	441	453
41	$u = (\infty 001 \times 010 \times 000 \times 0101111 \times 0011 \times 010000 \times 0000 \times 0000 \times 10000 \times 100000 \times 10000 \times 100000 \times 100000 \times 100000 \times 100000000$	453	465
42	$u = (\infty 0111 \times 100 \times 100 \times 101 \times 01000 \times 011 \times 1010011 \times 000 \times 10000)$ C ₁ : 42 × 35 matrix	465	471
46	$u = (\infty 00000 \times 1100010 \times 101 \times 10001 \times 1000110 \times 1000110 \times 1000000 \times 1100100000000$	477	483
47	$u = (\infty 0011 \times 1101 \times 10000 \times 1001 \times 000 \times 111010 \times 000 \times 10000 \times 100000 \times 100000 \times 100000 \times 100000 \times 100000 \times 100000000$	483	489
48	$u = (01 \infty \infty 11 \infty 01 \infty 1010111 \infty 001 \infty 000 \infty 000100 00000000$	489	495
51	$u = (\infty 0 \infty \infty 101011 \infty 000 \infty \infty 11 \infty 1 \infty 1001 \infty \infty \infty \infty$	501	507

Table 5: Improved strength four covering arrays for $g = 3$ (continued).						
k	Starter vectors and matrix C_1	New	Old			
		bound	bour			

		bound	bound
55	$u = (1 \infty 1 \infty 1 \infty 0 \infty 111 \infty 1 \infty 0010 \infty 00 0 \infty 0011011 \infty 1 \infty $	513	519
57	$u = (\infty 10 \infty 0011 \times 01 \times 10 \times 1001 \times 1001 \times 100000000$	519	531
58	$u = (\infty 0 \infty 00101 \infty 0010 \infty 0 \infty 1 \infty 1000 \infty 0 \infty $	525	531
63	$u = (1101 \times 10 \times 100 \times 0000000) (101 \times 0000 \times 0000000) (10001100000) (100000000000000$	537	549
67	$u = (010101 \times 1100 \times 100 \times 11 \times 0000110 \times 01111 \times 010110 \times 00000000$	555	561
70	$u = (1 \approx 0.01 \approx 11 \approx 1 \approx 0.000 \approx 11 \approx 0.000 \approx 1000011 \approx 0.0000000000$	567	573
72	$u = (\infty 000 \times 1010 \times 000 \times 010111000 \times 11011 \times 011101 \times 000001 \times 00000000$	573	579
74	$u = (1 \approx 0010 \approx 0010 \approx 0111 \approx 10000100 \approx 000100 \approx 10000100000000$	585	591

(g,k)	Vector v with good coverage	Our Results	Best known
		n (µ)	n [4]
(3,16)	00001001∞∞011∞1∞	99 (0.828)	237
(3,17)	0000010∞∞101∞01∞1	105 (0.851)	282
(3,18)	$00010\infty0\infty1001\infty111\infty\infty$	111 (0.864)	293
(3,19)	000010010∞01∞0∞111∞	117 (0.883)	305
(3,20)	$0000110101\infty0\infty10\infty\infty11\infty$	123 (0.892)	314
(3,21)	$00001010 \infty 1 \infty \infty 10 \infty \infty 001 \infty 1$	129 (0.906)	315
(3,22)	$0000011\infty0\infty0110\infty1\infty\infty001\infty$	135 (0.913)	315
(3,23)	$0000001\infty\infty0101\infty10\infty10\infty\infty\infty1$	141 (0.923)	315
(3,24)	$0000001 \infty 0101 \infty 10 \infty 101 \infty 1$	147 (0.924)	315
(3,25)	$000000011 \infty 0 \infty 011 \infty 0 1 \infty 0 \infty 11 \infty$	153 (0.930)	363
(3,28)	$1 \infty \infty 00 \infty \infty 1 \infty 01101111 \infty 0 \infty 0101 \infty \infty \infty 1$	171 (0.957)	383
(3,29)	$010\infty00\infty1\infty0\infty\infty\infty101\infty00\infty000111\infty10$	177 (0.961)	392
(3,30)	$011 \infty 11 \infty \infty 001 \infty \infty 1 \infty 10 \infty 0 \infty 1100 \infty 01$	163 (0.969)	393
(3,35)	$01 \\ \infty \\ 0 \\ \infty \\ \infty \\ 1000 \\ \infty \\ 01 \\ \infty \\ 0 \\ \infty \\ 0 \\ 1 \\ \infty \\ \infty \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0$	213 (0.979)	429
(3,36)	11∞0110∞∞00∞111101011∞001∞∞∞∞∞∞100∞0∞	219 (0.981)	441
(3,38)	$1 \infty 1 \infty 1 1 1 \infty \infty 0 10 \infty 10 \infty \infty 0 001 0 \infty \infty 0 \infty \infty \infty 1101 \infty \infty 100 \infty$	231 (0.985)	447
(3,39)	$001\infty\infty11\infty11\infty0001\infty11\infty101\infty\infty\infty1\infty0\infty0010\infty00\infty\infty0$	237 (0.986)	453
(3,40)	$100 \infty \infty 00001 \infty \infty 1 \infty 10 \infty 000 \infty \infty \infty 0 \infty 10 \infty \infty 1 \infty 0111 \infty 01$	243 (0.988)	465
(4,18)	00010021∞∞∞21020∞2	436 (0.851)	760
(4,19)	0000121011∞01∞0∞221	460 (0.866)	760
(4,20)	0000112101202∞0221∞2	484 (0.878)	760
(4,21)	0000011021010∞2∞0221∞	508 (0.887)	1012
(4,22)	0000001102∞02021∞∞01∞1	532 (0.894)	1012
(4,23)	00000001210210∞∞20112∞1	556 (0.898)	1012
(4,24)	00000000121∞011∞02∞0∞112	580 (0.899)	1012
(4,25)	00000000121220∞011∞2012∞	604 (0.901)	1012
(4,26)	00100∞2221110102∞0022∞020∞2	628 (0.921)	1012
(4,27)	0100∞2221110102∞0022∞020∞2	652 (0.928)	1012
(4,28)	01110∞0102∞021110022001∞1001	676 (0.933)	1012
(4,29)	0∞∞122101∞000220200221220∞02	702 (0.937)	1012
(4,30)	$10\infty20\infty020\infty2\infty2\infty01\infty2222\infty022002\infty1$	726 (0.943)	1012

Table 6: A comparison of the number of test cases (n) produced by our construction with high coverage measure and best known *n* for full coverage. For g = 5, the elements of GF(4) are represented as 0,1, 2, and 3; here 2 stands for *x* and 3 stands for x + 1.

Table 7: A comparison of the number of test cases (n) produced by our construction with high coverage measure and best known n for full coverage (continued).

initeu).			
(g,k)	Vector v with good coverage	Our Results	Best known
		n (µ)	n [4]
(5,21)	110131300∞30010∞∞3203	1265 (0.834)	1865
(5,22)	3∞32011200∞∞00∞0∞10010	1325 (0.842)	1865
(5,23)	0002∞03100∞203021332320	1385 (0.854)	1865
(5,24)	003∞21022212300032302310	1445 (0.860)	1865
(5,25)	∞200∞0∞∞31020∞300303∞∞33	1505 (0.869)	2485
(5,26)	202002211000∞0121031∞∞2300	1565 (0.873)	2485
(5,27)	∞∞03002030∞000∞11∞0031301∞3	1625 (0.880)	2485
(5,28)	013333130320∞1∞1003200310300	1685 (0.883)	2485
(5,29)	00012212∞010∞3110031020031010	1745 (0.891)	2485
(5,30)	33001∞0∞000330∞∞010012∞1313001	1805 (0.894)	2485
(5,31)	033∞21333010313∞303320030012020	1865 (0.895)	2485
(5,32)	310031000∞330130321∞∞03031111310	1925 (0.897)	2485
(5,33)	$\infty 0010 \infty \infty 3 \infty 0 \infty 2 \infty 01 \infty 00 \infty 12222 \infty \infty 03 \infty 020 \infty$	1985 (0.904)	2485
(5,34)	∞∞3∞00101001∞0∞001∞002∞01110231112	2045 (0.906)	2485
(5,35)	1203003303∞0∞013233310∞032020003220	2105 (0.906)	2485
(5,36)	12022∞3203230023223220001010200∞2230	2165 (0.912)	2485
(6,25)	000403014003033404320∞1∞∞	3006 (0.811)	6325
(6,26)	∞0∞40021404010013010011444	3126 (0.819)	6456
(6,27)	433∞∞01∞∞20∞03020∞∞0∞00401∞	3246 (0.826)	6606
(6,28)	4023031100232200∞21∞∞2020020	3366 (0.829)	6714
(6,29)	00∞40023103301343401230334400	3486 (0.834)	6852
(6,30)	1∞∞∞42∞4040004∞104∞03034∞∞0300	3606 (0.836)	6966
(6,31)	44122002∞2000020202031∞42044001	3726 (0.838)	7092
(6,32)	44441341∞424000∞∞040004410103400	3846 (0.846)	7200
(6,33)	0330344∞0232133100313000030∞4303∞	3966 (0.855)	7320